



Modèles statistiques pour l'estimation de la matrice fondamentale

Nicolas Noury, Frédéric Sur, Marie-Odile Berger

► To cite this version:

Nicolas Noury, Frédéric Sur, Marie-Odile Berger. Modèles statistiques pour l'estimation de la matrice fondamentale. Congrès francophone des jeunes chercheurs en vision par ordinateur - ORASIS'07, Jun 2007, Obernai, France. pp.8. inria-00164805

HAL Id: inria-00164805

<https://inria.hal.science/inria-00164805>

Submitted on 20 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Modèles statistiques pour l'estimation de la matrice fondamentale

Statistical models for fundamental matrix estimation

Nicolas Noury

Frédéric Sur

Marie-Odile Berger

INRIA Lorraine / INPL / LORIA

LORIA, Campus Scientifique - BP 239, 54506 Vandœuvre-lès-Nancy Cedex, France
{noury, sur, berger}@loria.fr

Résumé

L'estimation de la matrice fondamentale entre deux vues est un problème central dans les problèmes d'analyse de la structure et du mouvement dans les séquences vidéos. L'estimation est généralement effectuée en deux étapes : 1) identification des couples de points d'intérêt en correspondance dans les deux vues, et 2) filtrage robuste des appariements. La réussite de cette seconde étape dépend de la précision de la première étape, ainsi que de différents seuils numériques. Le réglage des seuils est délicat et empêche la réalisation de procédures automatisées d'évaluation de pose. L. Moisan et B. Stival [8] ont proposé un modèle statistique permettant de supprimer ces seuils. Nous montrons sur des données synthétiques et réelles l'intérêt de leur méthode en termes de robustesse, de précision et de temps de calcul. D'autre part, à la suite de leurs travaux, nous proposons un algorithme permettant simultanément d'apparier les points d'intérêt et d'estimer la matrice fondamentale entre deux vues. Nous montrons que cet algorithme est robuste dans le cas de motifs répétés et donc difficiles à apparier sans ambiguïté.

Mots Clef

Matrice fondamentale, modèle probabiliste, motifs répétés.

Abstract

Fundamental matrix estimation between two views is a cornerstone of structure from motion problems. Estimation is usually achieved in a twofold procedure : 1) identify matching points of interest between the two views, and 2) sort out the best matches through a robust filtering. The success of this latter step depends on the accuracy of the former one, and on several thresholds. Setting those thresholds is quite touchy and makes it difficult to automate the whole process. L. Moisan et B. Stival [8] have proposed a statistical model that enables to get rid of these thresholds. We assess over real and synthetic data that this model performs better than existing ones, especially from a robustness, accuracy, and computation time point of view. Besides, following their works, we propose an integrated algorithm that allows simultaneously to match interest points

and to estimate the fundamental matrix between two views. We show that this algorithm is robust toward repeated patterns which are difficult to unambiguously match.

Keywords

Fundamental matrix, probabilistic model, repeated patterns.

1 Introduction

1.1 Contexte

Le problème de l'analyse de la structure et du mouvement dans une séquence vidéo consiste à estimer la pose de la caméra et la structure des objets d'une scène. L'évaluation de la matrice fondamentale entre deux vues en est souvent une des premières étapes. Elle s'opère généralement à partir d'appariements obtenus par des méthodes d'extraction de points d'intérêt et de mise en correspondance (appariement) de ces points. La présence de données aberrantes dans les appariements détectés impose l'utilisation de méthodes robustes pour l'estimation de la matrice fondamentale. Néanmoins celles-ci nécessitent de fixer de nombreux seuils pour déterminer si un appariement donné est valide (et doit alors être pris en compte dans l'estimation) ou bien aberrant. Ces seuils sont délicats à régler car ils dépendent de la nature de la scène (netteté, illumination, dimensions des structures visibles, qualité des appariements fournis...).

1.2 Estimation de la matrice fondamentale

L'estimation de la matrice fondamentale entre deux vues est donc effectuée à partir de couples de points appariés d'une vue à l'autre. Comme cette matrice est une matrice carrée d'ordre 3 à 7 degrés de liberté, il est possible de déterminer la matrice fondamentale avec 8 couples de points (méthode de Longuet-Higgins [6]) : cette méthode produit un résultat unique mais requiert de normaliser les données pour des raisons de stabilité numérique et d'imposer la contrainte de rang à 2 en annulant la plus faible valeur propre. Il s'avère préférable d'utiliser la méthode dite « des 7 points » [10] qui ne nécessite pas de normalisation préalable, mais fournit (jusqu'à) trois matrices candidates. Le

critère utilisé joue un rôle important. Plutôt que minimiser un résidu algébrique, minimiser un critère tel qu'une somme de distances aux épipolaires [3], permet un gain en précision au détriment d'une complexité algorithmique plus élevée.

Bien sûr, calculer la matrice fondamentale à partir de seulement 7 ou 8 appariements pose des problèmes de stabilité lorsque ces appariements ne sont pas fiables. D'autre part, on dispose généralement d'un assez grand nombre de points appariés entre les deux vues considérées. C'est pourquoi on cherche des algorithmes permettant de sélectionner des collections d'appariements permettant une estimation fiable de la matrice fondamentale.

Une des approches les plus classiques est basée sur l'algorithme Ransac [4]. Cet algorithme comporte deux seuils principaux, l'un sur la distance au modèle (en dessous duquel on décide que les observations considérées sont des réalisations du modèle), l'autre sur la taille de l'ensemble dit de consensus (au dessus duquel on considère avoir trouvé un ensemble convenable pour estimer les paramètres du modèle). Ces deux seuils sont délicats à régler en pratique. Des approches statistiques tentent d'apporter des réponses à ce problème. Par exemple Torr [11] cherche à maximiser une vraisemblance plutôt que la taille de l'ensemble de consensus : plutôt que de réaliser une simple classification séparant les appariements aberrants de ceux fiables, il prend en compte les différences de précision existant parmi les appariements corrects. D'autres auteurs établissent un modèle *a priori* de répartition des appariements corrects [2], en réalisant une recherche exhaustive dans les images considérées. Moisan et Stival [8] proposent d'utiliser un modèle *a contrario* pour estimer la matrice fondamentale, ce qui a l'avantage de supprimer les seuils intervenant dans des approches type Ransac.

1.3 Contributions

Nos contributions consistent en deux apports : d'abord, nous avons validé la modélisation *a contrario* de [8] : des tests synthétiques mettent en évidence les gains en robustesse, précision et vitesse. Alors que dans l'article original de Moisan et Stival les appariements entre deux vues sont sélectionnés à la main, nous évaluons également dans des cas réalistes leur algorithme pour des appariements de points d'intérêt fournis par le détecteur Sift de D. Lowe [7]. Ensuite, nous étendons un algorithme esquissé dans [8] de manière à réaliser simultanément appariement de points d'intérêt et calcul de la matrice fondamentale. Alors que l'estimation de la matrice fondamentale se fait usuellement uniquement sur des listes de couples de points d'intérêt appariés dans une étape préliminaire, l'algorithme intégré permet de sélectionner les appariements parmi les multiples choix possibles de couples de points d'intérêt et de pondérer la contribution de ces appariements dans l'estimation de la matrice fondamentale, selon leur qualité.

L'apport de ces informations complémentaires nous permet d'obtenir de meilleurs résultats pour le calcul de la matrice

fondamentale dans des scènes contenant des motifs répétés comme des textures. Le nombre d'appariements corrects obtenus est en effet significativement augmenté, ainsi que la précision de l'estimation de la matrice fondamentale.

1.4 Plan de l'article

Après avoir rappelé les fondements du modèle *a contrario* dans la section 2, nous validons son apport par des tests sur des données synthétiques et dans le cas d'images réelles (section 3). Nous présentons le modèle *a contrario* permettant d'estimer simultanément appariements entre points d'intérêt et la matrice fondamentale dans la section 4, que nous validons expérimentalement dans la section 5.

2 Modèle *a contrario* pour l'estimation de la matrice fondamentale

Dans cette section nous présentons le modèle proposé par L. Moisan et B. Stival. Il s'agit d'une présentation synthétique, les preuves et détails d'implantation se trouvent dans [8].

2.1 Motivations

À partir d'un ensemble d'observations, l'algorithme Ransac permet de séparer les observations réalisations d'un modèle paramétrique (dont on cherche à déterminer les paramètres) des observations indépendantes de ce modèle, dues à une mise en correspondance erronée. Cet algorithme comporte deux seuils, l'un sur la distance au modèle (en dessous duquel on décide que les observations considérées sont réalisations du modèle), l'autre sur la taille de l'ensemble dit de consensus (au dessus duquel on considère avoir trouvé un ensemble suffisamment grand pour estimer les paramètres du modèle). Ces deux seuils sont délicats à régler en pratique. De plus on ne dispose pas de critère pour juger *a priori* de la qualité de l'ensemble de consensus : cette qualité dépend à la fois de la taille de l'ensemble de consensus et de l'adéquation des observations retenues par rapport au modèle.

Dans le cadre de l'estimation de la matrice fondamentale entre deux vues, L. Moisan et B. Stival [8] proposent une approche statistique qui ne nécessite pas de fixer de seuil arbitraire et fournit une mesure de la pertinence de la solution trouvée. L'approche est basée sur un modèle *a contrario* : un ensemble d'observations est d'autant plus significatif qu'il a peu de chances d'être produit « par hasard ». Les notions d'événement significatif et de modèle statistique *a contrario* ont été proposées par A. Desolneux, L. Moisan et J.-M. Morel (voir par exemple [1]).

2.2 α -rigidité

On cherche à évaluer la matrice fondamentale F entre deux images \mathcal{I}_1 et \mathcal{I}_2 à partir d'une collection S d'appariements (des couples (m, m') de points d'intérêt en correspondance, le premier dans \mathcal{I}_1 , le second dans \mathcal{I}_2). Comme un appariement (m, m') est correct au sens de la géométrie

épipolaire si m' est sur la droite Fm , on dispose d'un critère naturel mesurant la rigidité des données S par rapport au modèle F :

Définition 1. Soit F une matrice 3×3 de rang 2. La F -rigidité de l'ensemble S d'appariements est

$$\alpha_F(S) = \frac{2D'}{A'} \max_{(m,m') \in S} \text{dist}(m', Fm)$$

où A' est l'aire du domaine de la seconde image, D' son diamètre, et dist la distance euclidienne d'un point à une droite.

$\alpha_F(S)$ est la probabilité pour que les points m' « tombent » à une distance inférieure à $\max_{(m,m') \in S} \text{dist}(m', Fm)$ des droites Fm , lorsqu'ils sont générés par un processus uniforme dans l'image \mathcal{I}_2 (preuve de la prop. 1 dans [8]).

On définit alors :

Définition 2. Un ensemble d'appariements S (de taille $n \geq 8$) est α -rigide s'il existe une matrice fondamentale F estimée à partir d'un sous-ensemble de 7 appariements tel que $\alpha_F(S) \leq \alpha$.

Donc plus α est petit, moins un ensemble α -rigide peut être le « produit du hasard ». L'explication alternative (cet ensemble respecte la contrainte épipolaire) sera alors retenue.

2.3 Ensembles d'appariements significatifs

Néanmoins, l'ensemble des appariements entre les deux vues contient des faux appariements (car les algorithmes de détection des appariements entre points d'intérêt en fournissent toujours). On cherche donc en fait des sous-ensembles α -rigides de l'ensemble des appariements, qui sont formés des « bons » appariements.

Considérons un ensemble S de n appariements. Il y a $\binom{n}{k}$ sous-ensembles de S à k éléments, $\binom{n-7}{k}$ choix pour k (qui varie entre 8 et n), $\binom{k}{7}$ choix pour les 7 couples de points permettant d'estimer la matrice fondamentale, et chacun de ces 7-uplets fournit (au plus) 3 matrices fondamentales. Il y a donc $3(n-7)\binom{n}{k}\binom{k}{7}$ matrices fondamentales à tester lorsque l'on examine les sous-ensembles de S .

D'autre part, on peut montrer que la probabilité qu'un ensemble de k appariements soit α -rigide « par hasard » (i.e. sous hypothèse d'indépendance entre les points d'intérêts qui n'interviennent pas dans le calcul de la matrice fondamentale) est inférieure à α^{k-7} .

On déduit alors que l'espérance du nombre des sous-ensembles α -rigides de taille k dans S (sous hypothèse d'indépendance des points d'intérêt et d'uniformité de leur distribution) est inférieure à

$$\varepsilon_2(\alpha, n, k) := 3(n-7)\binom{n}{k}\binom{k}{7}\alpha^{k-7}.$$

Ceci motive la définition suivante (prop. 2 dans [8]).

Définition 3. Un ensemble $S' \subset S$ de k appariements parmi n est ε -significatif s'il est α -rigide et $\varepsilon_2(\alpha, n, k) \leq \varepsilon$.

La quantité $\varepsilon_2(\alpha, n, k)$ majorant le nombre de sous-ensembles comme S' attendus sous l'hypothèse d'indépendance des points d'intérêt, si ε_2 est faible (en particulier inférieur à 1) alors l'hypothèse d'indépendance est à rejeter et les appariements de S' respectent la contrainte épipolaire. On peut remarquer que ε_2 réalise un compromis entre la précision α des appariements selon la contrainte épipolaire et le nombre k de points dans l'ensemble considéré.

2.4 Algorithme

Comme rechercher le sous-ensemble le plus significatif parmi les $3(n-7)\binom{n}{k}\binom{k}{7}$ cas est impossible, une heuristique de type Ransac est proposée dans [8]. Dans cette heuristique, on cherche à minimiser la quantité ε_2 , qui remplace les deux paramètres de l'algorithme Ransac classique rappelés plus haut. D'autre part, une optimisation est mise en place :

- dans un premier temps, la recherche est effectuée sur la totalité des appariements ;
- ensuite, lorsque la valeur de ε_2 passe en dessous de 1 (moins d'une fausse alarme attendue), les échantillons sont tirés uniquement dans l'ensemble significatif correspondant, qui ne contient quasiment plus de faux appariements.

Cet algorithme ne contient plus de véritable paramètre, si ce n'est $\varepsilon_2 \leq 1$. Il est appelé Orsa dans [8] pour Optimized Random Sampling Algorithm.

3 Validation expérimentale

Pour valider la méthode Orsa, nous avons réalisé à la fois des tests synthétiques et des évaluations dans le cas réel. Les tests synthétiques ont permis d'évaluer les gains obtenus en robustesse, précision et vitesse par rapport aux approches Ransac classiques. Avec l'étude dans le cas réel, nous avons vérifié l'efficacité de la méthode à partir d'appariements extraits à partir des descripteurs Sift [7] : l'étude réalisée par Moisan et Stival avait porté uniquement sur des appariements déterminés à la main.

3.1 Tests synthétiques

Protocole expérimental. Le banc de test est réalisé en utilisant un corpus de 1400 appariements synthétiques : après un tirage uniforme des coordonnées 3D dans un intervalle donné, le point est projeté dans les deux plans images ; si chacun des points 2D trouvés est respectivement dans la fenêtre du plan constituant l'image, le point est conservé et les coordonnées 2D ainsi calculées permettent d'établir un nouvel appariement (appelé *inlier*).

On applique ensuite un bruit uniforme sur les coordonnées 2D, d'une amplitude fixée. On simule ainsi le bruit d'extraction des appariements.

Enfin, on remplace une proportion du corpus par des *outliers*, points dont les coordonnées sont tirées uniformément

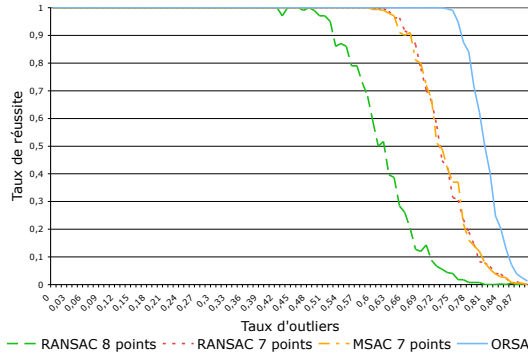


FIG. 1 – Taux de réussite dans le calcul de la matrice fondamentale en fonction du taux d'outliers dans la base d'appariements. L'algorithme Orsa présente une meilleure robustesse que l'évaluation de la matrice fondamentale par les algorithmes Msac et Ransac (que le modèle soit obtenu par la méthode des 7 points ou celle des 8 points).

sur le rectangle défini par les points d'abscisses et d'ordonnées minimales et maximales calculées lors de la génération des appariements.

Le calcul de la matrice fondamentale est effectué sur la première moitié des points, et une phase de réestimation non linéaire est effectuée après la sélection de l'ensemble d'inliers par les approches robustes. L'évaluation de la précision est réalisée sur la seconde moitié, en prenant en compte la moyenne des mesures de distance à l'épipolaire pour chaque appariement cohérent.

Test de robustesse. Le test est répété 100 fois pour chaque proportion d'outliers pour établir des taux de réussite. On fixe ici l'amplitude du bruit uniforme appliqué aux coordonnées à 1, ce qui correspond à une précision d'extraction plausible pour les points d'intérêts. L'évaluation de la robustesse est réalisée en calculant la distance moyenne à l'épipolaire pour les couples de points corrects de l'ensemble de validation. Si celle-ci est inférieure au bruit d'extraction maximum (ici égal à 1), on considère que le modèle (la matrice fondamentale) est correctement évalué. L'algorithme Orsa montre sa supériorité en obtenant un taux de réussite proche de 1 jusqu'à 80% d'outliers (seuil cependant un peu moins élevé que les 90% annoncés dans [8]) contre au mieux 65% pour un Ransac classique (cf. figure 1). Les seuils des algorithmes Ransac et Msac ont été réglés de manière à obtenir un meilleur comportement en moyenne. Rappelons que l'algorithme Orsa ne nécessite pas de seuil. Les algorithmes Ransac 7 points et Msac montrent leur supériorité sur l'approche utilisant un échantillon composé de 8 points : il est en effet nécessaire de considérer moins de cas pour estimer un modèle correct à partir de 7 points que pour 8.

Test de précision. La précision de l'estimation du modèle s'avère bien meilleure avec l'algorithme Orsa qu'avec les algorithmes Ransac dès que le taux t d'outliers dépasse 50% (cf. figure 2).

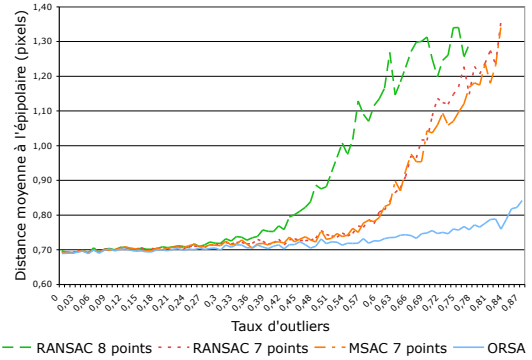


FIG. 2 – Distance moyenne à l'épipolaire en fonction du taux d'outliers. L'algorithme Orsa permet une précision bien meilleure que les algorithmes Ransac et Msac dès que le taux d'outliers dépasse 50%.

Tests de vitesse d'exécution. On réalise ici des moyennes sur 30 expériences pour chaque précision d'appariement et taux de points aberrants.

Pour l'algorithme Orsa, la vitesse d'exécution dépend essentiellement de la présence d'outliers (cf. figure 3), tandis que pour l'algorithme Ransac, la précision des appariements joue un rôle prépondérant (cf. figure 4). La phase d'optimisation réalise un nombre fixé d'itérations, sans test d'arrêt : le temps d'exécution minimal d'Orsa est donc plus long.

Par contre, pour Orsa, le temps d'exécution ne varie pas en fonction de la précision des appariements : c'est une conséquence de l'absence de seuils. Le paramètre ε_2 , qui inclut l'évaluation de la qualité du tirage, permet un arrêt de l'évaluation avec un ensemble de consensus plus petit car de précision garantie. La meilleure résistance aux outliers d'Orsa se paie par l'augmentation du temps de calcul pour des valeurs du taux d'outliers plus élevées que pour Ransac.

Le gain en vitesse global pour Orsa est dû à la phase d'optimisation. Le nombre de fausses alarmes, alors en dessous de 1, garantit l'existence d'une solution correcte et permet de ne considérer ensuite qu'un nombre réduit de cas.

3.2 Tests sur une séquence vidéo

Après cette étude sur des données synthétiques, on évalue maintenant le comportement de l'approche Orsa pour une séquence vidéo, en utilisant le détecteur Sift pour déterminer des points d'intérêt et les appariements entre ces points (voir figure 5). On réalise des tests sur une série d'images dont on connaît par ailleurs les paramètres des caméras : la vérité terrain permet de calculer la matrice fondamentale théorique. Il est alors possible d'évaluer la qualité des appariements sélectionnés en calculant la distance à l'épipolaire pour cette matrice fondamentale. On compare les moyennes de cette distance réalisée sur tous les inliers, pour chaque couple d'images.

La figure 6 présente les précisions des modèles calculés en fonction du numéro du couple d'images considéré. Le

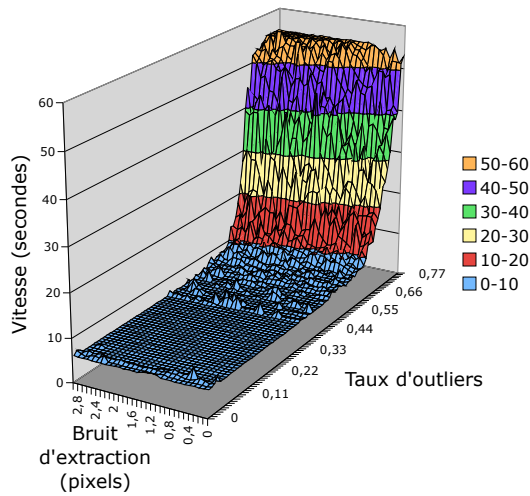


FIG. 3 – Temps d'exécution de l'algorithme Orsa en fonction du bruit d'extraction sur les points d'intérêt et de la proportion d'outliers. Le temps d'exécution est stable par rapport au bruit, et augmente avec le taux d'outliers. Ceci s'explique par le plus grand nombre d'itérations alors nécessaire avant de trouver un ensemble 1-significatif qui permet de rentrer dans la phase d'optimisation.

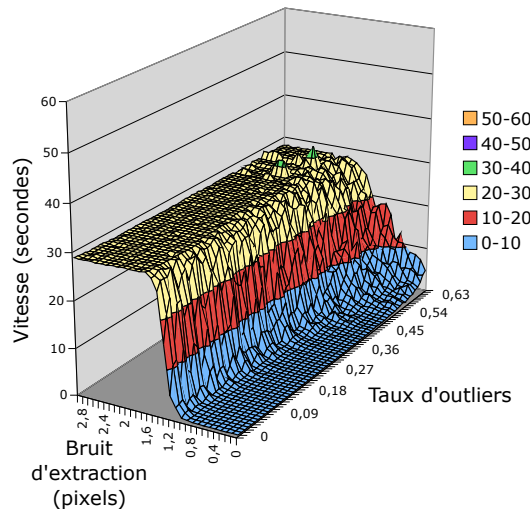


FIG. 4 – Temps d'exécution de l'algorithme Ransac en fonction du bruit d'extraction sur les points d'intérêt et de la proportion d'outliers. Le temps d'exécution augmente légèrement avec le taux d'outliers mais dépend fortement de la précision sur les points d'intérêt. On observe un effet de seuil lorsque le bruit devient relativement plus élevé que le seuil sur la distance au modèle au dessus duquel les échantillons sont rejetés de l'ensemble de consensus.

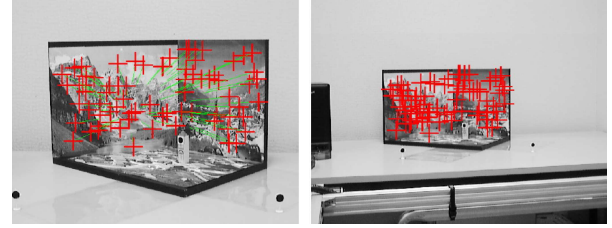


FIG. 5 – Exemple d'appariements obtenus pour un des couples d'images de la séquence. On estime la matrice fondamentale sur des couples séparés par un intervalle de 50 images.

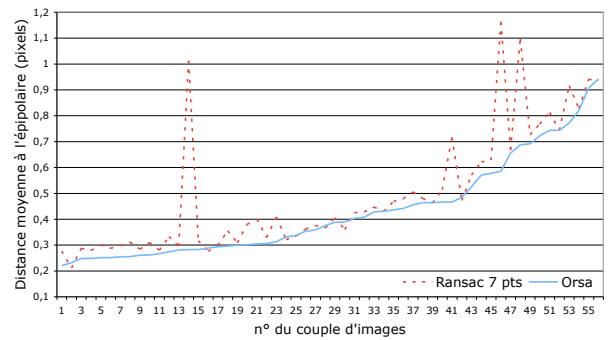


FIG. 6 – Précision obtenue pour la séquence vidéo : l'algorithme Orsa donne systématiquement des résultats meilleurs que l'algorithme Ransac 7 points. Les pics indiquent les cas où les seuils fixés pour Ransac ne correspondent pas à la précision d'extraction.

graphique a été réalisé en classant les couples d'images en fonction de la précision obtenue par le modèle *a contrario*. On obtient une confirmation des résultats synthétiques : au mieux, la précision de la méthode Ransac est celle de l'approche Orsa. Les seuils choisis pour Ransac permettent un bon comportement dans la plupart des cas, mais donnent une évaluation du modèle très mauvaise dans certains cas. Encore une fois, Orsa n'impose de régler aucun seuil.

4 Appariement et sélection simultanée des appariements

Généralement, l'estimation de la matrice fondamentale se fait en deux étapes : une première étape d'identification de couples de points d'intérêt en correspondance, et une deuxième étape d'estimation de la matrice fondamentale proprement dite. La première étape peut être basée sur la comparaison de voisinages locaux (des *patches*) des points d'intérêt dans l'image en niveaux de gris. L'information qui a servi à prendre la décision d'appariement (une distance, une corrélation, etc) n'est alors plus prise en compte dans la deuxième étape. Cette approche présente deux inconvénients : les mauvaises décisions d'appariement sont définitives et ne peuvent plus être rattrapées dans la deuxième étape, et on ne pondère pas les appariements selon la « qua-

lité » de leur ressemblance. De plus, dans le cas d'une scène possédant des motifs répétitifs (textures régulières, fenêtres sur des façades, pavages...) la mise en correspondance ne peut pas se faire uniquement sur un critère local mais doit tenir compte des contraintes géométriques.

Nous adaptons un modèle *a contrario* pour traiter les deux étapes simultanément. Il s'agit d'une adaptation de ce qui est esquissé dans [8] sous le terme « colored rigidity ».

4.1 Modélisation

On suppose à présent ne disposer que des listes de points d'intérêt dans les images, les appariements étant inconnus. Nous avons vu en section 2.3 qu'un ensemble de k appariements est α -rigide sous hypothèse d'indépendance à la probabilité α^{k-7} . Introduisons la notion de c -rigidité. Supposons que les points d'intérêt sont munis d'une (pseudo-)distance $\delta(des(m), des(m'))$ qui mesure la similitude entre des descripteurs locaux $des(m)$ et $des(m')$ autour de m et m' .

Définition 4. Un ensemble S de k appariements est c -rigide si

$$\max_{(m,m') \in S} \delta(des(m), des(m')) \leq c.$$

Avec cette définition, la probabilité pour que S soit c -rigide, sous hypothèse d'indépendance des points, est inférieure à c^k , et donc la probabilité pour que S soit (α, c) -rigide, sous hypothèse d'indépendance entre points d'intérêt et entre patches locaux, est inférieure à $\alpha^{k-7} P(c)^k$, où $P(c) = \Pr(\max_{(m,m') \in S} \delta(des(m), des(m')) \leq c)$. Ceci motive la définition suivante :

Définition 5. Un ensemble S' de k appariements est ε -significatif s'il est (α, c) -rigide avec

$$\varepsilon_3(\alpha, c, n, k) := 3(n-7) \binom{n}{k} \alpha^{k-7} c^k \leq \varepsilon,$$

où ici n est le nombre de tous les appariements considérés entre les deux images.

Bien sûr il ne s'agit pas de la même notion de significativité que dans la définition 3.

4.2 Choix d'implantation

Plusieurs éléments vont être ici spécifiés : 1) les patches locaux utilisés, la distance choisie, et la distribution de probabilité $P(c)$, 2) les appariements potentiels effectivement considérés (il n'est pas question de lister tous les appariements possibles entre tous les points d'intérêt extraits), 3) quelle heuristique utiliser pour trouver un ensemble d'appariements minimisant ε_3 .

Patches locaux. Dans cet article nous avons utilisé l'algorithme Sift de D. Lowe [7] comme extracteur de points d'intérêt et descripteur de patches locaux. Sift n'étant qu'invariant aux similitudes, on supposera que les transformations géométriques sont suffisamment bien approchées

par des similitudes localement. La ressemblance entre ces descripteurs est difficile à mesurer. Déterminer une distance entre deux descripteurs Sift revient à calculer une distance entre histogrammes. Ce sujet est étudié dans [5], et il ressort qu'un compromis entre vitesse d'évaluation et fiabilité est obtenu avec la distance du χ^2 , qui sera utilisée ici. Une fois la distance choisie, nous évaluons la probabilité $P(c)$ en considérant l'histogramme des fréquences empiriques cumulées, estimé sur l'ensemble de tous les appariements possibles.

Appariements à considérer. L'étude de la répartition des fréquences empiriques des distances entre patches met en évidence que les appariements corrects correspondent aux paires de points ayant une faible distance devant la majorité des autres. Cette caractéristique permet de ne considérer qu'un sous-ensemble des candidats à l'appariement : on peut éliminer les autres *a priori*. Cette coupure (qui reste très grossière : l'essentiel de la sélection n'est pas fait ici) réduit la combinatoire du problème.

Algorithme. Nous recherchons les groupes d'appariements significatifs selon l'algorithme suivant, qui reprend les principes d'Orsa.

On introduit alors la notion de *groupe de candidats à l'appariement*, groupe associant à un point d'intérêt de la première image, tous les points de la seconde image non éliminés par la coupure sur les distances entre descripteurs. Pour sélectionner un appariement dans un groupe de candidats, on utilise deux approches. Dans le cas du tirage d'un échantillon sans connaissance sur le groupe, l'appariement est choisi au hasard. Par contre, pour la construction de l'ensemble (α, c) -rigide, on sélectionne dans le groupe l'appariement ayant la distance épipolaire minimale. L'algorithme s'écrit :

- extraire les points d'intérêts (Sift)
- calculer la distribution $P(c)$
- construire les *groupes de candidats à l'appariement*
- soit $\bar{\varepsilon} = +\infty$, répéter
 1. tirer un échantillon : un ensemble T de 7 groupes, et un appariement par groupe
 2. évaluer l'échantillon en construisant l'ensemble (α, c) -rigide correspondant. Pour chaque matrice fondamentale F associée à T :
 - calculer α pour chaque groupe
 - classer les groupes par α croissant
 - calculer le maximum des $P(c)$ des appariements pour chacun des ensembles de taille et d' α croissants formés
 - déterminer l'ensemble le plus significatif $U = \bar{S}(F)$ associé à F
 3. si $\varepsilon_3(U) < \bar{\varepsilon}$, alors $\bar{\varepsilon} = \varepsilon_3(U)$ et $\bar{S} = U$
- jusqu'à ce que le nombre d'essais soit supérieur à N
- renvoyer \bar{S} et $\bar{\varepsilon}$.

Une phase d'optimisation est ensuite appliquée dès que la significativité atteint le seuil universel de 1 ; l'échantillon utilisé est alors tiré uniquement parmi les appariements du précédent ensemble trouvé, et non parmi tous les groupes.

Pour chaque groupe choisi pour l'évaluation de l'échantillon, on reprend le précédent appariement sélectionné au lieu de le tirer au sort parmi les candidats du groupe.

5 Résultats

L'information apportée par la prise en compte des patches locaux permet une sélection plus fine des appariements, en particulier en présence de motifs répétés. En effet, la technique d'appariement habituellement utilisée, qui repose sur la recherche de l'appariement de distance minimale au sens de la distance entre descripteur Sift, impose de plus que la distance entre patches locaux pour l'appariement retenu soit suffisamment différente de l'appariement possible suivant. Cette contrainte défavorise donc en particulier les points d'intérêt extraits sur des motifs répétés. Dans ce cas, plusieurs appariements impliquant un même point possèdent des distances très proches : n'ayant pas assez de connaissance pour résoudre l'ambiguïté, aucun n'est retenu.

Notre technique permet d'utiliser la connaissance obtenue sur ces appariements ambigus. S'il n'était pas possible de choisir un appariement parmi les candidats en utilisant uniquement la proximité entre patches, l'utilisation combinée de la contrainte géométrique et de la ressemblance entre patches permet d'opérer une telle sélection.

La figure 7 permet de comparer le nombre d'appariements produit dans le cas d'une séquence synthétique, produite avec Blender. L'approche présentée en section 4 permet d'extraire le double d'appariement (193 contre 97), améliorant ainsi la qualité de l'estimation (On passe de 0.7 pixels de distance moyenne à l'épipolaire à 0.2 pour l'approche combinée contre plus de 1 pour les méthodes Ransac classique).

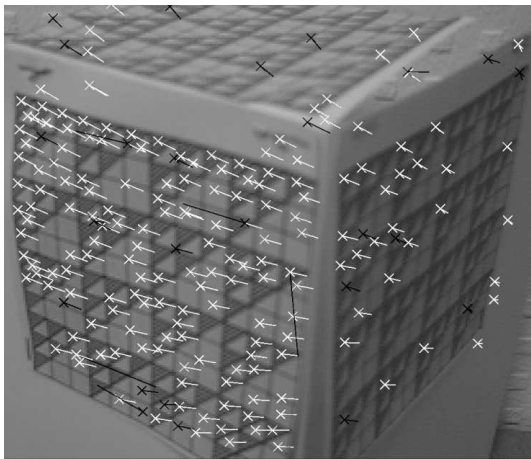


FIG. 8 – Exemple d'appariements dans le cas de motifs répétés avec Orsa. Pour le côté face, les répétitions réduisent le nombre d'appariements retenus par cette méthode.

Les figures 8 et 9 sont obtenues à partir d'une vidéo 320×240 acquise via un appareil photo numérique. La figure 8 montre les appariements retenus lors de l'approche

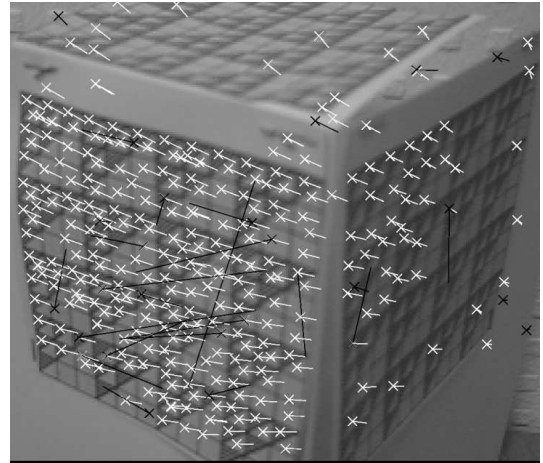


FIG. 9 – Avec la prise en compte des patches locaux : deux fois plus d'appariements fiables sont générés et leur utilisation permet une meilleure estimation de la matrice fondamentale.

Orsa (section 2). Dans la figure 9, on obtient plus d'appariements, les points d'intérêt ambigus n'étant plus systématiquement rejetés. Cela permet d'augmenter d'un facteur important (ici presque 2) le nombre d'appariements extraits dans les zones correspondants aux motifs répétés. La qualité de la matrice fondamentale obtenue est au pire équivalente, et meilleure quand les appariements retenus initialement ne sont pas assez nombreux pour permettre une estimation fiable par l'algorithme Orsa.

6 Conclusion et perspectives

Dans cet article nous avons présenté une validation du modèle *a contrario* pour l'estimation de la matrice fondamentale entre deux vues. Alors que dans l'article original de L. Moisan et B. Stival cette validation est menée uniquement sur des points d'intérêt appariés à la main, nous menons des tests systématiques sur des données synthétiques et des données réelles à l'aide d'appariements fournis par le détecteur Sift de D. Lowe. Nous montrons que cette méthode sans paramètre permet des gains en précision et robustesse par rapport aux approches type Ransac, que seuls des analyses et réglages empiriques dépendants des conditions d'observation (luminosité, netteté) permettaient jusqu'alors.

Nous avons également proposé un algorithme permettant de réaliser appariement de descripteurs locaux et estimation de la matrice fondamentale simultanément, par un modèle *a contrario* intégrant les conditions géométriques et de ressemblances locales des images. Cette approche permet de renforcer les contraintes sur les appariements, qui doivent à la fois respecter une ressemblance perceptuelle et la géométrie épipolaire. La précision de l'estimation de la matrice fondamentale est semblable à celle obtenue par la méthode précédente, mais davantage de couples de points d'intérêt sont détectés comme conformes à la géométrie

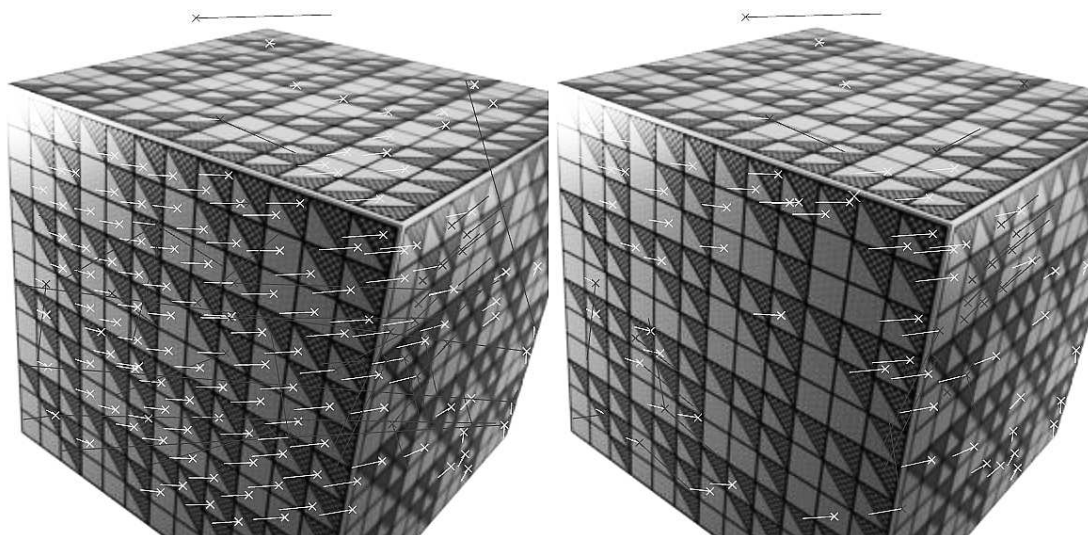


FIG. 7 – Cas synthétique : on obtient deux fois plus d'appariements avec l'approche de la section 4 (à gauche) que pour Orsa (section 2); la matrice fondamentale estimée est alors plus précise (0.2 pixels de distance moyenne à l'épipoilaire contre 0.7). Les croix représentent les points d'intérêts et les traits qu'elles terminent représentent le mouvement apparent de l'appariement. Ils sont en blanc s'ils sont considérés corrects, et en noir si aberrants.

épipoilaire. De plus, ceci permet de résoudre le problème d'estimation de la matrice fondamentale dans le cas de scènes présentant des structures répétées. Contrairement à J. Domke [2], nous n'abandonnons pas l'utilisation des points d'intérêts. Son approche permet aussi la gestion des répétitions mais avec un coût en temps de calcul prohibitif, car estimant une probabilité de ressemblance pour tous les pixels de l'image. Les méthodes d'extraction de points d'intérêt allègent l'analyse en évitant de considérer tous les points de l'image lors de l'estimation de la matrice fondamentale.

Néanmoins, la qualité des résultats reste tributaire de l'invariance des descripteurs locaux. Les descripteurs Sift n'étant qu'invariant aux similitudes, d'autres descripteurs sont envisagés. Il est à noter que notre algorithme nécessite seulement une métrique pour comparer les descripteurs mais est indépendant de la manière dont ces descripteurs sont construits. D'autre part, pour s'approcher d'une application temps-réel, des optimisations comme des tests préemptifs [9] sont à étudier si l'on souhaite intégrer cette approche aux techniques de suivi.

Références

- [1] A. Desolneux, L. Moisan, and J.-M. Morel. Maximal Meaningful Events and Applications to Image Analysis. *Annals of Statistics*, 31(6), 2003.
- [2] J. Domke and Y. Aloimonos. A Probabilistic Notion of Correspondence and the Epipolar Constraint. In *Proc. 3DPVT (International Symposium on 3D Data Processing Visualization and Transmission)*, 2006.
- [3] O. Faugeras, Q.-T. Luong, and S. Maybank. Camera Self-Calibration : Theory and Experiments. In *Proceedings of 2nd European Conference on Computer Vision, Santa Margherita Ligure (Italy)*, pages 321–334, mai 1992.
- [4] M. Fischler and R. Bolles. Random Sample Consensus : A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6) :381–395, 1981.
- [5] H. Ling and K. Okada. Diffusion distance for histogram comparison. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 246–253, 2006.
- [6] H. Longuet-Higgins. A computer program for reconstructing a scene from two projections. *Nature*, 293 :133–135, septembre 1981.
- [7] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2) :91–110, 2004.
- [8] L. Moisan and B. Stival. A Probabilistic Criterion to Detect Rigid Point Matches between Two Images and Estimate the Fundamental Matrix. *International Journal of Computer Vision*, 57(3) :201–218, 2004.
- [9] D. Nistér. Preemptive RANSAC for Live Structure and Motion Estimation. *Machine Vision and Applications*, 16(5) :321–329, December 2005.
- [10] P. Torr and D. Murray. Outlier Detection and Motion Segmentation. In P. S. Schenker, editor, *Sensor Fusion VI*, pages 432–443. SPIE volume 2059, 1993. Boston.
- [11] P. Torr and A. Zisserman. MLESAC : A new robust estimator with application to estimating image geometry. *Computer Vision and Image Understanding*, 78 :138–156, 2000.